Validation of Simulation, With and Without Real Data

Kleijnen, J.P.C.

Publication date:
1998

A:\VALDATA.WPD

Printed: June 15, 1998 (12:22PM)

Written: March 5, 1998

Validation of Simulation, With and Without Real Data

JACK P.C. KLEIJNEN

Department of Information Systems and Auditing (BIKA)/Center for Economic Research (CentER)

Tilburg University (Katholieke Universiteit Brabant), 5000 LE Tilburg, Netherlands

e-mail: kleijnen@kub.nl

fax: +3113-4663377

web: http://cwis.kub.nl/~few5/center/staff/kleijnen/cv2.htm

Version 1: February 1998

Validation of Simulation, With and Without Real Data

JACK P.C. KLEIJNEN

Department of Information Systems and Auditing (BIKA)/Center for Economic Research (CentER)

Tilburg University (Katholieke Universiteit Brabant), 5000 LE Tilburg, Netherlands

e-mail: kleijnen@kub.nl; fax: +3113-4663377

web: http://cwis.kub.nl/~few5/center/staff/kleijnen/cv2.htm

Abstract

This paper gives a survey on how to validate simulation models through the application of mathematical statistics. The type of statistical test actually applied, depends on the availability of data on the real system: (i) no data, (ii) only output data, and (iii) both input and output data. In case (i), the system analysts can still experiment with the simulation model to obtain simulated data. Those experiments should be guided by the statistical theory on design of experiments (DOE); an inferior - but popular - approach is to change only one factor at a time. In case (ii), real and simulated output data may be compared through the well-known Student t statistic. In case (iii), trace-driven simulation becomes possible. Then validation, however, should not proceed as follows: make a scatter plot with real and simulated outputs, fit a line, and test whether that line has unit slope and passes through the origin. Instead, better tests are presented. Several case studies are summarized, to illustrate the three types of situations.

(Keywords: verification, credibility, assessment, sensitivity, robustness, regression)

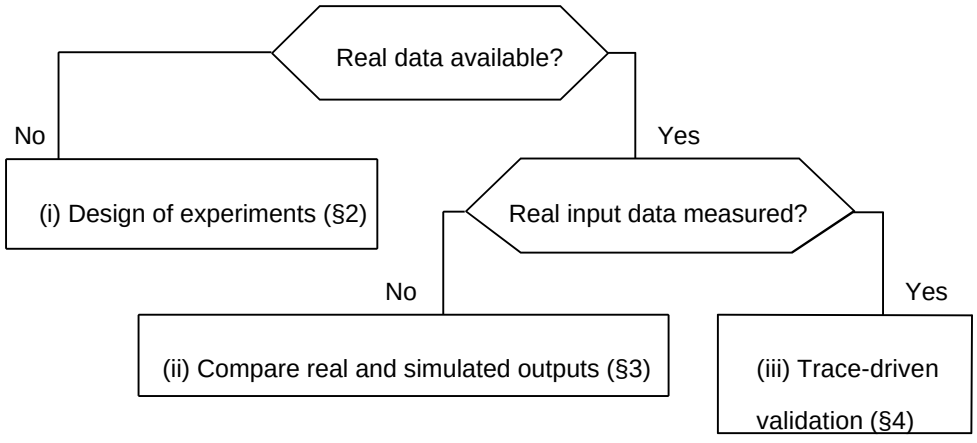
1. Introduction

What do I mean by ‘validation’ and ‘real data’? For this paper it suffices to define *validation* as determining whether the conceptual simulation model is an accurate representation of the real system (also see the classic textbook on simulation, Law and Kelton 1991).

Many types of validation are used in practice, but I shall focus on validation that uses *mathematical statistics*. After all, simulation means experimentation (albeit with a model instead of the real system). And experimentation calls for statistical analysis, preceded by statistical design, as I shall show. Obviously, statistical analysis is only part of the whole validation process (other parts are graphical summaries, the Schruben-Turing test on ‘face validity’, etc.). However, if statistics are used, then the correct statistics should be used!

One type of statistical validation compares data on the real and the simulated systems. Such a comparison makes much more sense if both systems are observed under similar scenarios; for example, a busy day at the real supermarket should not be compared with a slow day at the simulated store. Obviously, real data may pertain to input and output; for example, customers’ arrival times and cashiers’ service times at the supermarket are input, whereas customers’ waiting times are output. I now discuss this in more detail, using Figure 1.

Figure 1: Real data availability: three situations



(i) In some applications, data on the real system are either completely missing or scarce. Examples are data on nuclear war (fortunately, no data, except for outdated figures on Hiroshima and Nagasaki), nuclear accidents (limited data: Chernobyl, Three Miles Island), global warming or greenhouse effect (few data; see Kleijnen, Van Ham, and Rotmans 1992).

But, even if real data are missing, there is still *expert knowledge*. For example, we all are experts in waiting at supermarkets, so we know that if more customers arrive per hour, then waiting times increase - unless more cashiers become active. However, this knowledge is qualitative; to obtain quantitative knowledge, the system analysts develop a simulation model. In other words, the sign of the effect is known, not its magnitude! If the simulation model's input/output (I/O) behavior violates this qualitative knowledge, the model should be seriously questioned: are there programming and conceptual errors? I shall propose a systematic, scientific method - namely, design of experiments or DOE - for selecting conditions or scenarios as input for the simulation model; and I shall present examples of simulation errors detected in this way (see next section, §2).

(ii) In other applications, however, we are 'drown by the numbers'; examples are data on supermarket sales and on telecommunication operations. In general, data are abundant if systems are electronically monitored; examples are point of sale systems (POSS) and electronic data interchange (EDI).

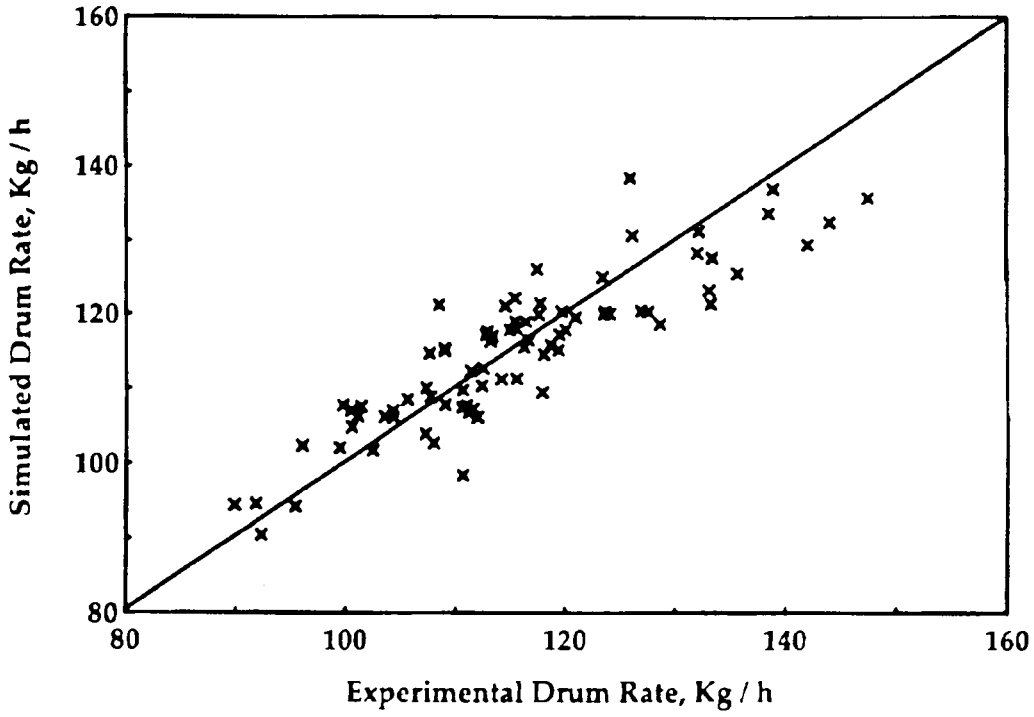
But, even if real output data are available, the corresponding real-world scenarios may not be measured. An example is a simulation model of the search for mines on the sea bottom, through sonar equipment: in practice it is virtually impossible to measure the temperature and the salinity of the sea water, at all times and places; see Kleijnen (1995a). In those cases only the outputs of the real and the simulated systems can be compared. I shall return to this case study in §3 and the appendix. (Notice that to obtain real data, the military conduct field tests and companies build pilot plants.)

(iii) Finally, the most powerful validation is possible if both input and output of the real system are measured.

In so-called *trace-driven* simulation, analysts feed real input data into the simulation program, in historical order (also see Law and Kelton 1991, p. 316). After running the simulation program, analysts compare the time series of simulated output with the historical time series of real output.

But, what is wrong with the following analysis of trace-driven simulation? Make a scatter plot with real and simulated outputs, fit a line to the scatter plot, test whether that line has unit (45°) slope and passes through the origin (zero intercept): see Figure 2. This figure is taken from the case study in Kozempel, Tomasula, and Craig (1995). Another example is Lysyk (1989). The latter author indeed performs a statistical test, and finds an estimated slope significantly smaller than unity and an intercept significantly positive. Since he expects a unit slope and a zero intercept, he tries to explain this phenomenon away. I shall answer this question in §4.

Figure 2: Wrong analysis of trace-driven simulation: an example (source: Kozempel et al. 1995, p. 232)



The literature on validation is abundant: see the web (http://manta.cs.vt.edu/biblio/) and the detailed survey in Kleijnen (1995b). In this literature, however, the focus in not on the role of data availability in the choice of statistical tests.

2. No Real Data Available: DOE

If no data on the real system are available, analysts can still perform *what-if* analysis: what happens if a particular simulation 'factor' changes? I use the DOE term 'factor' to denote a parameter, an input variable, or a module of a simulation model. In the supermarket example, parameters are the arrival and service rates; an input variable is the number of cashiers; a module may be the submodel for the priority rules (First-In-First-Out or FIFO, priority for customers with less than - say - ten items).

A simulation experiment consists of a set of simulation runs. During a simulation run, all factors remain constant. During the experiment, factors do change; that is, each factor has at least two levels or 'values' in the experiment. The factor may be *qualitative*, as the priority rules exemplified. A detailed discussion of qualitative factors and various measurement scales is given in Kleijnen (1987, pp. 138-142).

Most practitioners change *one factor at a time*, and think that this is *the* scientific way to perform what-if analysis (even Pirsig 1974 seems to think so). Actually it is easy to prove that this method gives less accurate estimates of the factor's (main or first-order) effect. Moreover, it does not enable estimation of interactions among factors: what happens if two or more factors change simultaneously? DOE provides much better estimators! But, what's DOE precisely? I base my answer to this question (i.e., the remainder of this section) on Kleijnen (1998).

DOE was started in the 1930s in agriculture; a popular example of the resulting designs is the class of so-called 2^{k-p} designs. DOE's central problem is how to select a limited set of combinations of factor levels to be observed, from the large number of conceivable combinations. An example is the ecological case study with 281 parameters in Bettonvil and Kleijnen (1997): at least 2^{281} ($> 10^{84}$) combinations may be distinguished. An example

with fewer (less than, say, fifteen) factors is the supermarket. In a simulation context, I define DOE as selecting the combinations of factor levels that will be actually simulated when experimenting with the simulation model. To illustrate this problem, I present a case study in the appendix.

After this selection of input combinations, the simulation program is executed or 'run'. Next DOE analyzes the resulting I/O data of the simulation experiment. One goal is to derive conclusions about the importance of the factors; in simulation this is also known as sensitivity analysis, which is related to what-if analysis, optimization, and validation. Unfortunately, there is no standard definition of sensitivity analysis. I define *sensitivity analysis* as the systematic investigation of the reaction of the simulation responses to *extreme* values of the model's input or to *drastic* changes in the model's structure. For example, what happens to the customers' mean waiting time when their arrival rate doubles; what happens if the priority rule is changed by introducing 'fast lanes'?

For this analysis, DOE uses *regression analysis*, also known as Analysis Of Variance or ANOVA. This is based on a *metamodel*, which is a model of the underlying simulation model (see Friedman 1996, Kleijnen 1987). In other words, a metamodel is an approximation of the simulation program's I/O transformation (the metamodel is also called a response surface). Typically, this model uses one of the following three polynomial approximations.

(i) a first-order polynomial, which consists of an overall or grand mean β_0 and k main effects (say) β_j with $j = 1, \ldots, k$ where k denotes the number of factors;

(ii) the same polynomial augmented with interactions between pairs of factors (two-factor interactions) $\beta_{j;j'}$ with $j' = j + 1, \ldots, k$;

(iii) a second-order polynomial, which adds purely quadratic effects $\beta_{j;j}$.

Notice that a first-degree polynomial misses interactions and has constant marginal effects; a third-order polynomial would be more difficult to interpret and would need many more simulation runs to estimate the many effects. So a second-order polynomial may be a good compromise, depending on the goal of the metamodel. The validation of this metamodel (not the underlying simulation model, which is the focus of this paper) may use the well-known

multiple correlation coefficient R^2. An example is given in the appendix. More refined tests (such as cross-validation and Rao's F test) are given in Kleijnen and Sargent (1997).

DOE and its regression analysis treat the simulation model as a *black box*: the simulation model's I/O is observed, and the factor effects in the metamodel are estimated. This approach has advantages and disadvantages. An advantage is that DOE can be applied to all simulation models, either deterministic or stochastic. A disadvantage is that DOE cannot exploit the specific structure of a given simulation model.

DOE is a classic topic in statistics. However, the standard statistical techniques must be adapted such that they account for the following *simulation peculiarities*.

(i) There are a great many factors in many practical simulation models. For example, the ecological case study (mentioned above) has 281 factors, whereas standard DOE assumes only up to (say) fifteen factors. 'Screening' aims at finding a short list of really important factors; again see Bettonvil and Kleijnen (1997).

(ii) Stochastic simulation models use pseudorandom numbers, which means that analysts have much more control over the noise in their experiments than they have in standard statistical applications. For example, to reduce that noise, analysts may use so-called common and antithetic numbers.

(iii) Randomization is of major concern in DOE outside simulation. In simulation, however, this randomization problem disappears: pseudorandom numbers take over.

The regression metamodel shows which factors are most important; that is, which factors have highly significant regression estimates in the metamodel. If possible, information on these factors should be collected, for validation purposes. (If the significant factors are controllable by the users, then the estimated regression effects show how to change these factors to optimize the real system; see Kleijnen and Pala 1998 for an application.)

DOE assumes that the area of experimentation is given. A valid simulation model, however, requires that the inputs be restricted to a certain domain of factor combinations. This domain corresponds with the *experimental frame* in Zeigler (1976), a seminal book on modeling and simulation.

Simulation models are often used in *risk analysis*: what is the probability of a 'disaster'? That disaster

may be a nuclear accident, a financial mis-investment, etc. I emphasize that these disasters are unique events, whereas the supermarket simulation concerns repetitive events (e.g., customer waiting times). Consequently, validation in risk analysis is very difficult. A better term may be *credibility*; also see Fossett, Harrison, Weintrob, and Gass (1991).

A technical issue in risk analysis is that DOE would select extreme combinations of factor values, which typically have extremely low probability of realization. Instead, risk analysis samples from the whole domain of possible combinations, according to a prespecified (joint) probability distribution. This sampling uses the Monte Carlo technique (sometimes refined to Latin hypercube sampling or LHS). Next, analysts improve the risk model's credibility by applying statistical techniques. For example, regression analysis and contingency tables may detect which factors have significant effects; next analysts - using their expert knowledge - should be able to explain why these factors are important. An example is the case study on nuclear waste disposal in the waste-isolation pilot-plant (WIPP) near Carlsbad, New Mexico (NM), USA. A model was developed at Sandia National Laboratories (SNL) in Albuquerque (NM). The Environmental Protection Agency (EPA) will give permission to start using the WIPP, only if the WIPP simulation model is accepted as credible - and the model's output shows an acceptable risk. See Helton, Anderson, Marietta, and Rechard (1997) and Kleijnen and Helton (1998).

The importance of sensitivity analysis in validation is also emphasized by Fossett et al. (1991); they present three military case studies. Another case study that does explicitly demonstrate the role of DOE and regression analysis in validation, is the ecological simulation in Bettonvil and Kleijnen (1997) and Kleijnen, Van Ham, and Rotmans (1992). The regression metamodel in Kleijnen et al.(1992, p. 415) helped to detect a serious error in the simulation model: one of the original modules should be split into two modules. Both publications further show that some factors are more important than the ecological experts originally expected. This 'surprise' gives more insight into the simulation model. I present another application in the appendix.

3. Real Output Data Available: Classic Statistical Tests

In the Introduction (§1) I mentioned that even if real output data are available, the corresponding real-world scenarios may not be measured. As an example I mentioned the sonar search, in which the Sound Velocity Profile (SVP) was not measured; the real output - namely the detection of mines - is measured. This real output can be compared with the final output of the total simulation model (for the intermediate outputs of the individual modules I refer to the appendix).

Let's first return to the familiar supermarket example. Suppose that the real and simulated outputs x and y are the average waiting time of the T customers served per day: $x = \sum_{t=1}^{T} w_t/T$ and $y = \sum_{t=1}^{T} v_t/T$. Suppose further that n days are simulated and m days are observed in the real system. Assume days give i.i.d. observations (no seasonality; only busy Saturdays simulated and measured). Define $\mu_d = \mu_x - \mu_y$ The n and m observations give the classic estimators $\bar{x}$, $\bar{y}$, s_x^2, and s_y^2 of the means and variances. This yields *Student's t statistic* with $n + m - 2$ degrees of freedom:

$$t_{n+m-2} = \frac{(\bar{x} - \bar{y}) - \mu_d}{[(n-1)s_x^2 + (m-1)s_y^2]^{1/2}} \frac{[(n+m-2)nm}{(n+m)^{1/2}} \qquad (1)$$

Obviously, the (null) hypothesis is that simulated and real average waiting times per day are equal; that is, H_0: $\mu_d = 0$. The power of this test increases, as in (1) μ_d increases (bigger differences are easier to detect), n or m increases (more days measured), or σ_x or σ_y decreases (less noise: more customers per day or lower traffic rate). Because $\sigma_d^2 = \sigma_x^2 + \sigma_y^2 - 2cov(x, y)$ analysts may try to create a positive covariance (or correlation) through the use of trace-driven simulation: see the next section (§4). Notice that a difference such as $\bar{x} - \bar{y}$ may be non-significant and yet important: if only a few days are simulated or there is much noise, then an important difference μ_d may go undetected. The reverse is also possible.

The sonar case study (mentioned before) gives a *binary* variable: detect or miss a mine. The (say) n simulation runs give a binomial variable with parameters n and p, the detection probability. Analogously, the field

test gives a binomial variable with parameters *m* and *q*. To test the (null) hypothesis of equal simulated and real probabilities (H_0: $p = q$), Kleijnen (1995a) uses the t-statistic as an approximate test.

Unfortunately, the t test assumes normally, independently, and identically distributed (n.i.i.d.) outputs. Simulation models usually give non-normal, autocorrelated, possibly non-stationary outputs. A simple solution is available if the simulation is *terminating*. Then each simulation run gives independently and identically distributed (i.i.d.) outputs. The t statistic is known to be not very sensitive to nonnormality.

Besides the t test, *distribution-free* tests (such as the rank test) may be applied; see Conover (1971). In practice, however, these tests are rare applied - unfortunately.

In *non-terminating* simulation, analysts may try to create i.i.d. observations through the batching or subrun approach; see Kleijnen (1987), Law and Kelton (1991).

4. Real I/O Data Available: Trace-driven Simulation

In the Introduction (§1) I claimed that it is wrong to analyze a trace-driven simulation by making a scatter plot with real and simulated outputs (say) *x* and *y*, fit a line $y = \beta_0 + \beta_1 x$, and test whether $\beta_1 = 1$ and $\beta_0 = 0$; again see Figure 2. Now let ρ_{xy} denote the classic linear correlation coefficient between *x* and *y*. Suppose the real and the simulated outputs have equal positive means: $\mu_x = \mu_y = \mu > 0$. Then it is easy to prove that imperfect correlation ($\rho_{xy} < 1$) gives $0 < \beta_1 < 1$ and $0 < \beta_0 < \mu$. So the naive regression analysis of the trace-driven simulation is wrong indeed.

Kleijnen, Bettonvil, and Van Groenendaal (1996) propose the following test. Compute not only the *n* differences d_i (also see equation 1 with $n = m$), but also the *n* sums (say) $q_i = x_i + y_i$. Now make a scatter plot with these differences and sums, fit a line $d = \gamma_0 + \gamma_1 q$, and test H_0: $\gamma_0 = 0$ and $\gamma_1 = 0$. Obviously, this (joint, composite) hypothesis implies $\mu_d = 0$ or $\mu_x = \mu_y$. Moreover, assuming normality for *x* and *y*, it is easy to prove that $\gamma_1 = 0$ implies equal variances: $\sigma_x^2 = \sigma_y^2$. To test the joint hypothesis,, analysts can use standard regression

software (which uses an F test).

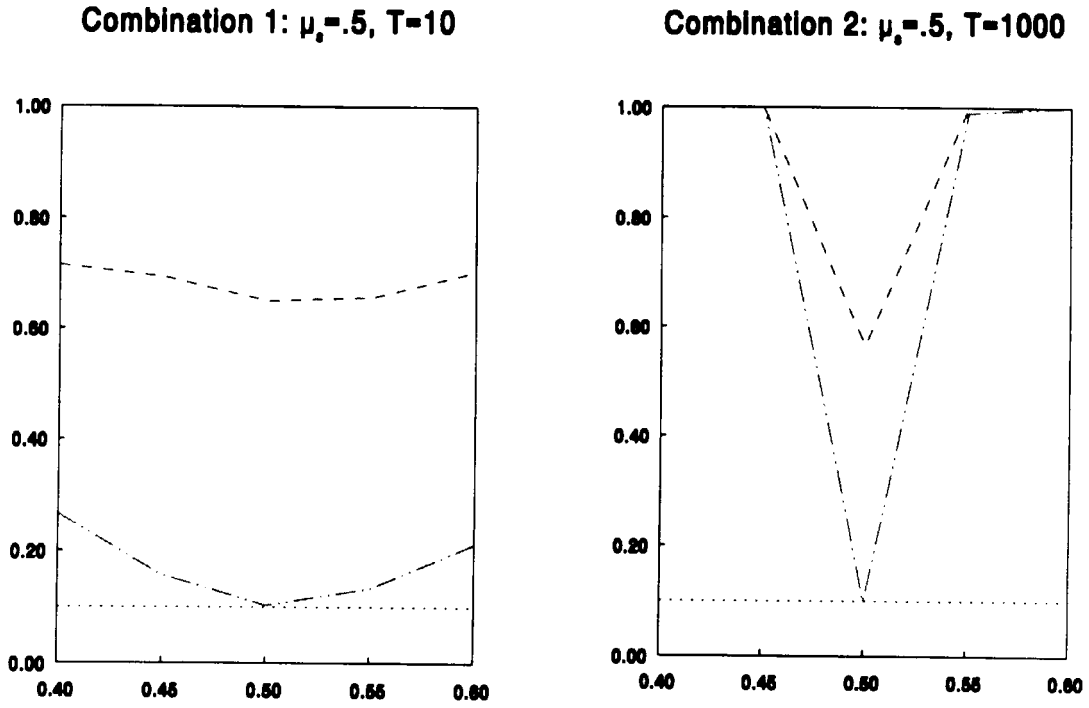
Kleijnen et al. (1996) evaluate both the naive and the novel regression analyses, applying them to single server systems with Poisson arrival and service times (M/M/1); these systems are terminating, since each day stops after *T* customers (jobs). This gives the following conclusions; also see Figure 3 where MST denotes mean service times:

(i) the naive test rejects a valid simulation model substantially more often than the novel test does;

(ii) the naive test shows perverse behavior in a certain domain: the worse the simulation model, the higher its probability of acceptance, and

(iii) the novel test does not reject a valid simulation model too often, provided the outputs are transformed logarithmically to realize normality.

Figure 3: Estimated power of naive (----) and novel (-.-.) tests for logarithmic transformed real and simulated outputs *x* and *y*, with varying $\tilde{\mu}_s$ (simulated MST) and fixed $\mu_s = 0.5$ (real MST), for varying *T* (jobs per day), given $n = 10$ (days), and $\alpha = 0.10$ (type I error rate) (Source: Kleijnen et al. 1996, Figure 2)



These regression analyses assume n.i.i.d. real and simulated outputs (as did the t test in the preceding section). Currently Kleijnen, Cheng, and Bettonvil (1998) are developing a test for non-normal, non-stationary (transient), autocorrelated observations. That new test uses bootstrapping, which is a type of Monte Carlo simulation; see Efron and Tibshirani (1993).

5. Conclusions

In practice, validation has many forms, but I focussed on validation through mathematical statistics. Such validation gives quantitative information on the quality of the simulation model (other types of validation - such as animation - give only 'face' validity).

Statistical validation may use various testing procedures, depending on the type of data available for the real system. I distinguished the following three situations.

(i) *No* real data

Then analysts can still generate simulated data. Their simulation experiment should be guided by DOE; an inferior approach changes only one factor at a time. Regression models provide approximations (metamodels) of the simulation's I/O transformation, and show which factors are important.

(ii) Only data on real *output*

Real and simulated outputs may be compared through the Student t test. Alternatives are distribution-free procedures, which - unfortunately - are applied rarely.

(iii) *I/O* data on real system

Real I/O data enable trace-driven simulation. The validation of trace-driven simulation, however, should not use a scatter plot with real and simulated outputs, fit a line, and test whether that line has unit slope and zero intercept. Instead, two alternatives were discussed: alternative #1 regresses sums and differences; it applies if the outputs are n.i.i.d. Alternative #2 applies bootstrapping; this alternative is still under construction.

I referenced several case studies, to demonstrate the applicability of the various statistical methods. Nevertheless, I think that validation will remain an art!

Appendix. Case Study: Mine Hunting on the High Seas

Explosives on the sea bottom may be detected (hunted) by means of sonar. A simulation model called HUNTOP (mine HUNTing OPerations) was developed for the Dutch navy, by Applied Scientific Research/Physics and Electronics Laboratory (TNO/FEL) in the Netherlands.

Kleijnen (1995a) validates HUNTOP in two stages. In stage #1 individual modules are validated. (In stage #2 the total simulation model is treated as one black box, and is validated; that stage is discussed in §3.) Some of these modules give *intermediate* output that is hard to observe in practice, and hence is hard to validate. Therefore sensitivity analysis is applied to these modules: check if factor effects have signs or directions that agree with experts' prior qualitative knowledge. For example, deeper water gives a wider sonar window; see the main effect β_2 in the sonar window module below. Because of time constraints, only the following two modules are validated; I use the symbol z for the original (non-standardized) factor values.

(i) Sonar window module

The sonar rays hit the bottom under an angle determined deterministically by three factors, namely z_1 or SVP that maps sound velocity as a function of depth, z_2 or average water depth, and z_3 or tilt angle. SVP is treated as a qualitative factor.

The sonar window module's output is y, the minimum distance of the area on the sea bottom 'insonified' by the sonar beam. Consider a set of second-degree polynomials in the two quantitative factors z_2 and z_3, namely one polynomial for each SVP type z_1. To estimate the six parameters of this polynomial, Kleijnen (1995a) uses the classical *central composite design* for two factors, which has nine input combinations; see Table 1 for the standardized values where + denotes +1, - denotes -1, $c = \sqrt{k}$ (k denotes the number of factors; here $k = 2$). (Other

values for *c* can be found in the literature; I propose $\sqrt{k}$ because the distance of the combinations with all factors at the absolute value +1, to the origin is $\sqrt{k}$; a simulation model is valid only within its experimental frame.)

Table 1: Central composite designs for two factors in standardized values

(Notation: + denotes +1, - denotes -1, $c = \sqrt{k}$ with k the number of factors)

Combination:	1	2	3	4	5	6	7	8	9
Factor 1:	+	-	+	-	c	-c	0	0	0
Factor 2:	-	+	-	+	0	0	c	-c	0

The fitted polynomial turns out to give an acceptable ('valid') approximation: the multiple correlation coefficient R^2 ranges between 0.96 and 0.98, for the four SVPs simulated. Expert knowledge suggests that certain factor effects have specific signs, namely $\beta_2 > 0$, $\beta_3 < 0$, and $\beta_{2,3} < 0$. Fortunately, the corresponding estimates turn out to have the correct signs. So this module has the correct I/O transformation, and the validity of this module need not be questioned.

Note that the quadratic effects turned out to be non-significant. So on hindsight, simulation runs could have been saved, since a smaller design (with only the first four runs in Table 1) would have sufficed.

(ii) Visibility module

An object is visible if it is within the sonar window, and it is not concealed by the bottom profile. The output is the time that the object is visible, expressed as a percentage of the time it would have been visible were the bottom flat. Kleijnen (1995a) varies six inputs.

Again Kleijnen (1995a) fits a quadratic polynomial, and uses a central composite design. Now, however, the polynomial has 28 regression parameters, and the design has 77 input combinations. It turns out that R^2 is 0.86. Further, the factor 'upward hill slope' has no significant effects at all: no main effect, no interactions with the other

factors, no quadratic effect. These results agree with the experts' qualitative knowledge. So the validity of this module is not questioned either.

I emphasize that central composite designs require many simulation runs. If the computer budget is tight, then alternative designs may be constructed. For example, Kleijnen and Pala (1998) derive a *saturated* design, which is a design with the number of runs equal to the number of factor effects to be estimated.

Acknowledgment

This paper is based on my talk at the Methodologists Day, 14 November 1997, Free University (VU), Amsterdam, organized by NOSMO (Netherlands Organization for Social Sciences Methodological Research), including the Section on Simulation; also see http://www.fsw.ruu.nl/ms/cvd.htm

References

Bettonvil, B. and J.P.C. Kleijnen (1997) Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research*, 96, no. 1, pp. 180-194

Conover, W.J. (1971), *Practical Non-parametric Statistics*. Wiley, New York

Efron, B. and R.J. Tibshirani (1993), *Introduction to the Bootstrap*. Chapman & Hall, London

Fossett, C.A., Harrison D., Weintrob H., and Gass S.I. (1991), An assessment procedure for simulation models: a case study, *Operations Research* 39, pp. 710-723

Friedman, L.W. (1996), *The simulation metamodel*. Kluwer, Dordrecht, Netherlands

Helton, J.C., D.R. Anderson, M.G. Marietta, and R.P. Rechard (1997), Performance assessment for the waste isolation pilot plant: from regulation to calculation for 40 CFR 191.13. *Operations Research*, 45, no. 2, pp. 157-177

Kleijnen, J.P.C. (1998), Experimental design for sensitivity analysis, optimization, and validation of simulation models. *Handbook of Simulation*, Jerry Banks, Editor, Wiley, New York

--- (1995a), Case study: statistical validation of simulation models. *European Journal of Operational Research*, 87, no. 1, pp. 21-34

--- (1995b), Verification and validation of simulation models. *European Journal of Operational Research*, 82, no. 1, April 1995, pp. 145-16

--- (1987) *Statistical tools for simulation practitioners*. New York: Marcel Dekker

---, B. Bettonvil, and W. Van Groenendaal (1996). Validation of trace-driven simulation models: a novel regression test. Discussion Paper, CentER, no. 9607. (*Management Science*, accepted 1996)

---, R.C.H. Cheng, and B. Bettonvil (1998),Validation of trace-driven simulation models: bootstrapped tests. Working Paper (in preparation)

--- and J. Helton (1998), Statistical analysis of scatter plots to identify important factors in large-scale simulations.

Sandia National Laboratories, Albuquerque, New Mexico

--- and Ő. Pala (1998), Maximizing the simulation output: a competition. CentER Discussion Paper

--- and R.G. Sargent (1997), A methodology for the fitting and validation of metamodels in simulation. CentER Discussion Paper, no. 97116

---, G. Van Ham, and J. Rotmans (1992), Techniques for sensitivity analysis of simulation models: a case study of the CO2 greenhouse effect. *Simulation*, 58, no. 6, pp. 410-417

Kozempel, M.F., Tomasula, P. and Craig, J.C. (1995), 'The development of the ERRC food process simulator', *Simulation; Practice and Theory*, 2, 4-5

Law A.M., and Kelton W.D. (1991), *Simulation Modeling and Analysis; Second Edition*, McGraw-Hill, New York

Lysyk, T.J. (1989), 'Stochastic model of Eastern spruce budworm (lepidoptera: tortricidae) phenology on white spruce and balsam fir', *Journal of Economic Entomology*, 82, 4, 1161-1168

Pirsig, R.M. (1974), *Zen and the art of motorcycle maintenance; an inquiry into values*. The Bodley Head, Ltd., London

Zeigler, B. (1976) *Theory of modelling and simulation.* New York: Wiley Interscience

Jack P.C. Kleijnen is Professor of Simulation and Information Systems. His research interests are in simulation, mathematical statistics, information systems, and logistics. He published six books and nearly 150 articles, was a consultant to several organizations in the USA and Europe, and was on many editorial boards and scientific committees. He spent a few years in the USA, at universities and companies. A number of international fellowships and prizes were awarded to him. For more information:

web: http://cwis.kub.nl/~few5/center/staff/kleijnen/cv2.htm